Metabolic Networks: How to Identify Key Components in the Regulation of Metabolism and Growth[1]

Mark Stitt*, Ronan Sulpice, and Joost Keurentjes

Max Planck Institute of Molecular Plant Physiology, 14476 Potsdam-Golm, Germany (M.S., R.S.); Laboratories of Genetics and Plant Physiology, Wageningen University, NL–6708 PB Wageningen, The Netherlands (J.K.); and Centre for Biosystems Genomics, NL–6708 PB Wageningen, The Netherlands (J.K.)

Plants display enormous diversity in their metabolism. Although the biosynthesis and function of the myriad plant metabolites have been studied for decades, we have little understanding of the interactions between metabolites, metabolite signaling, interactions with development, and the role of metabolism in genotype-to-phenotype relationships. Technologies for the analysis of metabolites have made tremendous progress in recent years, both in terms of the number of metabolites that are identified and of throughput. Recent developments allow the construction of metabolic networks and study of the role of these networks in plant growth and development. In this review, we discuss what types of information can be obtained from measurements of metabolites and what requirements they have with respect to the comprehensiveness of coverage and the precision of identification and quantification, and we outline procedures that can be implemented to validate the measurements. We then discuss what sorts of perturbations can be used to disturb metabolic networks, including environmental and physiological treatments, chemicals, reverse genetics, and the use of natural genetic diversity.

Plants are the most consummate and sophisticated chemical system in the world. They use light energy to convert CO_2 into carbohydrates in their leaves. They absorb nutrients like nitrate, phosphate, and sulfate via their roots and convert them to amino acids and nucleotides, using light energy in the leaves in the day and energy derived from respiration in leaves in the dark and in nonphotosynthetic tissues. Carbohydrates, amino acids, and nucleotides are then transported to growing tissues, where they are converted into macromolecular cellular components like proteins, nucleic acids, cell walls, pigments, and lipids. Plants also synthesize tens of thousands of secondary metabolites, including phenylpropanoids and flavonoids, terpenoids, glucosinolates, and alkaloids. These have important roles in cellular function, in signaling, and in adaptation to abiotic and biotic stress. Their unique synthetic ability is the result of a highly complex and sophisticated metabolic apparatus. Its complexity and flexibility were already appreciated by the 1980s, as biochemical studies of cellular compartmentation revealed that many basic pathways like glycolysis, the oxidative pentose phosphate pathway, and organic acid metabolism were present in more than one compartment (Lunn, 2007), and the diversity of plant secondary metabolites was unveiled. It was further underlined as genome sequencing uncovered families of genes for enzymes in central metabolism, including starch and Suc synthesis and degradation, glycolysis, many reactions of the tricarboxylic acid cycle and nucleotide degradation, and huge families of genes that encode enzymes that introduce modifications into basic metabolite structures, like the alcohol dehydrogenases, 2-oxoglutarate dioxygenases, acyl transferases, UDP glucosyltransferases, O-methyl transferases, nitrilases, cytochrome P50s, myrosinases, and oxidases (Kaul et al., 2000; http://mapman.gabipd.org/web/guest/mapman).

Traditionally, metabolism has been divided into discrete pathways. However, it has become increasingly clear that metabolism operates as a highly integrated network (Sweetlove et al., 2008). Synthesis of one metabolite typically requires the operation of many pathways. Often, the accumulation of a metabolite depends on its rate of degradation as well as the rate of synthesis. Metabolites are not synthesized in isolation from each other; rather, large sets of metabolites must often be synthesized simultaneously, for example, diverse lipids and pigments during the formation of the photosynthetic thylakoid membranes. Synthesis of one set of metabolites often requires that resources be diverted away from the synthesis of other sets of metabolites. More generally, metabolites can only be synthesized if carbon, nitrogen, phosphor, and sulfur and the basic building blocks generated from them in central metabolism are available. This implies that regulatory networks gear metabolic activities to the availability of these basic resources. Discussions about the regulation of metabolic flux have moved,

with time, from an emphasis on a small number of "key" regulatory or "rate-limiting" enzymes (Rolleston, 1972) to the realization that changes in flux often require changes in the activities of multiple enzymes in a metabolic sequence (Kacser and Burns, 1973; Kacser and Acerenza, 1993; Fell and Thomas, 1995).

Analysis of metabolic networks at a systems level depends on the existence of a series of technologies and conceptual advances that support (1) a comprehensive analysis of the levels of metabolites, (2) analysis of fluxes, (3) suitable methods to perturb the metabolic system, (4) computational methods to process, store, evaluate, and integrate the resulting data, and (5) the development of kinetic and other models to synthesize and conceptualize the data. This emerging research field is often termed "metabolomics," playing on the well-established jargon for large-scale analyses of transcript and protein levels. There have been many reviews of different aspects of metabolomics in the last years. They provide an overview of the available technology for analysis of metabolite levels (Kopka et al., 2004; Hall, 2006; Lisec et al., 2006; Last et al., 2007), the use of metabolic flux analysis to analyze fluxes in complex metabolic systems (Fernie et al., 2005; Noh et al., 2006; Antoniewicz et al., 2007; Libourel and Shachar-Hill, 2008; Kruger and Ratcliffe, 2009), reporting standards (Goodacre et al., 2007; Taylor et al., 2008), and statistical and modeling methods (Liao et al., 2003; Fernie et al., 2004; Janes and Yaffe, 2006; Goodacre et al., 2007; Guy et al., 2008; Steinfath et al., 2008). This update focuses on a set of nontechnical questions that are also of central importance for any systems analysis of plant metabolism. (1) What kinds of information can be obtained from evaluation of large metabolite data sets? (2) What are the demands on data quality and validation (see also the excellent review by Last et al. [2007])? (3) What complications are introduced by spatial and temporal heterogeneities in plant metabolism? (4) What sort of interventions can be used to perturb plant metabolic systems?

WHY MEASURE METABOLITES? AND HOW GOOD DO THE DATA NEED TO BE?

The Parts List

At the simplest level, measurements of metabolites provide a description of the capacity and scope of the system: what is there? Such data sets provide information about changes in the spectrum of metabolites in different tissues or conditions in a given genotype or how they change between different genotypes or species. The advent of broad untargeted analytic platforms has greatly increased the power of this approach. It is not essential that all the metabolites be identified or quantified. However, to avoid a proliferation of "stand-alone" studies, it is important that peaks are identified using criteria and procedures that can be accessed and checked (e.g. retention index, parent ion mass, and

fragment ion masses; Goodacre et al., 2007; Taylor et al., 2008).

Another interesting use of a metabolite compendium is to support genome annotation. Evidently, the genome must contain all the enzymes required to synthesize (and, in general, to degrade) a given metabolite. Briefly, the list of identified metabolites is compared with a draft metabolic pathway generated from the list of genes predicted by genome annotation This comparison highlights errors and provides a powerful method to assist and improve genome annotation (May et al., 2009). However, routine use of public domain databases will require improvements in biochemical databases. As highlighted by Kind et al. (2009), while sequence comparisons allow in silico integration of information in genome sequence, transcriptomics, and proteomics databases, additional experimental evidence is often needed to define the precise catalytic activity of enzymes. The development of publishing standards and database architectures that aid the reconstruction of metabolomes and metabolic networks is an important future task. It will be an especially large challenge for plant secondary metabolism, which shows enormous interspecific and intraspecific variation.

Correlation and Multivariate Analysis

Metabolite data sets are often used in combination with statistical analysis to generate hypotheses about functional relationships, using metabolite data sets on their own or in combination with other types of data, such as expression profiles or proteomics data (Hirai et al., 2004, 2005; Oresic et al., 2004; Gibon et al., 2006; Rischer et al., 2006; Goodacre et al., 2007; Kusano et al., 2007; Bylesjö et al., 2009; Lehmann et al., 2009). This requires reliable relative data but does not require absolute quantification. While a broad spectrum of metabolites provides information about general dependencies, for focused questions it is often important to have comprehensive coverage of a defined area of metabolism (Sønderby et al., 2007).

At the simplest level, correlation networks can be used to identify which components might be functionally related, based on the "guilt by association" principle (Yonekura-Sakakibara et al., 2008). In complex multifactorial systems, additional information can be extracted using more sophisticated statistical methods, including principal components analysis, canonical correlation analysis, partial least squares regression, and mutual information (Janes and Yaffe, 2006; Steinfath et al., 2008). For example, in a recombinant inbred line (RIL) Arabidopsis (*Arabidopsis thaliana*) population, many low-M_r metabolites correlated negatively with biomass, but the correlations were not significant (Meyer et al., 2007). Deeper insights were provided by applying canonical correlation analysis and partial least squares regression. These multivariate methods identify which combination and weighting of a large set of predictors (in this case,

about 150 metabolites) provide the best prediction of a target trait (in this case, biomass; Steinfath et al., 2008). These authors identified a set of metabolites that provided a highly significant prediction of biomass. In a subsequent study, Sulpice et al. (2009) used an analogous approach with 92 Arabidopsis accessions. They found that the same set of metabolites (and with the same weighting) correlates negatively with biomass, with the starch content at the end of the light period, and with the protein content. Starch accumulates in the light and is remobilized at night to support metabolism and growth. Protein is a major component of a plant cell, and its synthesis and maintenance require considerable energy. This analysis prompted the hypothesis that fast-growing accessions utilize their starch reserves more efficiently for growth and that decreased synthesis of protein is one of the factors that contributes to this increase in the efficiency of use of carbon.

Bioinformatics tools are being refined to extract even more information from multilayered "omics" data sets. One example is O2PLS (for Orthogonal Projections to Latent Structures; Trygg and Wold, 2002, 2003; Bylesjö et al., 2008), which allow systemic predictive variation that links different types of parameters to be separated from parameter-specific variation, which is generated by the various platforms that are used to measure them. Another important area is the development of approaches that allow the use of time-series data for modeling and prediction (Smilde et al., 2005b; Rantalainen et al., 2008).

Computational Approaches to Deal with Missing Information

A deeper understanding of a metabolic network requires information about its overall structure and the properties of its individual components. The latter includes information about the activities and kinetic parameters of enzymes, the absolute concentrations of metabolites, and fluxes. The structure of pathways (i.e. the sequences of reactions) is usually known in primary metabolism and well-studied secondary pathways but will continue to be an important research goal in specialized secondary pathways. A dearth of reliable data about enzyme parameters will continue to be a bugbear for attempts to model metabolism.

Consequently, the development of theoretical approaches that allow metabolic pathways to be analyzed and modeled with incomplete information is one of the most rapidly developing and important areas in metabolomics (Schilling et al., 2002; Liao et al., 2003; Reed and Palsson, 2004; Noh et al., 2006; Antoniewicz et al., 2007; Grafahrend-Belau et al., 2009).

The power of such approaches is illustrated by the recent reconstruction of an Arabidopsis metabolic network with 1,406 reactions (Poolman et al., 2009). The model was built from the known reactions in the database AraCyc (www.arabidopsis.org) and (see below) experimental data about the (ir)reversibility of the reactions. Linear programming was then used to

show that the network is capable of producing amino acids, nucleotides, starch, cellulose, and lipids in the ratio needed for growth of heterotrophic cell cultures. Interestingly, the reactions could be grouped into modules based on their predicted responses to changes in the ATP supply. However, this study also illustrates a limitation of "information-sparse" approaches. The model predicted that the carboxylase reaction of Rubisco operates to maximize growth at low ATP supplies while the oxygenase reaction operates at higher ATP supplies. As the authors discuss, prior knowledge that was not integrated into the model makes this a rather unlikely scenario. The relative rates of the carboxylase and oxygenase reactions depend on the kinetic properties of Rubisco, and it is unlikely that plants possess a regulatory mechanism that allows these reactions to vary independently of each other. In this specific case, the nonfeasibility is obvious because the necessary information is available. In other cases, this may be less evident. That said, predictions are always valuable, provided they act as a trigger for detailed functional analysis to validate or invalidate the predictions.

In parallel, it is important to continue to improve the quality of the experimental data with respect to scope and quantification to provide better parameterization and constraints to models. This includes information about the catalytic and kinetic properties of enzymes and about the concentrations of enzyme-binding sites and of metabolites that act as ligands at these sites.

Thermodynamic Organization of Metabolic Systems

The architecture of a metabolic pathway is not defined solely by the reaction sequence and the kinetic parameters of enzymes. A further important aspect is the thermodynamic organization. This has implications for modeling the energetic efficiency of metabolic transformations. It also affects which experimental and theoretical strategies are taken to study metabolic regulation.

The change of free energy in vivo (ΔG) can be calculated from the standard free energy of the reaction ($\Delta G°'$) and the molar concentrations of the substrates ($[S_1] \ldots [S_i]$) and products ($[P_1] \ldots [P_i]$), using the equation $\Delta G = \Delta G°' - RT \ln\{([S_1]*[S_i])/([P_1]*[P_i])\}$. This calculation requires precise quantitative measurements of metabolites. In the case of reactions that have unequal numbers of substrates and products (e.g. aldolase), precise estimates of the molar concentrations are needed. The term $RT \ln\{([S_1]*[S_i])/([P_1]*[P_i])\}$ is very sensitive to changes in the absolute concentrations because most metabolites are present at millimolar or micromolar concentrations.

Some reactions are far removed from equilibrium in vivo (i.e. they proceed with a major loss of free energy), making them effectively "irreversible" (Rolleston, 1972). For others, the in vivo concentrations of substrates and products are such that the reaction is close to equilibrium. In this case, the net direction can be

changed by small changes in the substrate-product ratio For example, in a classic study of the Calvin-Benson cycle in *Chlorella*, Bassham and Krause (1969) showed that Fru-1,6-bisphosphatase, seduheptulose-1,7-bisphosphatase, phosphoribulokinase, and Rubisco catalyze irreversible reactions, while all the other enzymes catalyze reactions that are close to equilibrium. In glycolysis, the only irreversible reactions are those catalyzed by hexokinase, phosphofructokinase, and pyruvate kinase (or phospho*enol*pyruvate carboxylase). Irreversible reaction steps can only be reversed via a different reaction (e.g. when the flux direction switches from glycolysis to gluconeogenesis, ATP-phosphofructokinase is inactivated and is replaced by Fru-1,6-bisphosphatase).

Regulation of Key Enzymes Versus Coordinated Regulation at Many Enzymes

There has been a long debate about whether the enzymes that catalyze irreversible reactions are the key sites for the regulation of metabolism. There is strong experimental evidence that they are often the targets of feedback and feed-forward loops, involving product inhibition and/or allosteric or posttranslational regulation. Accordingly, it has been suggested that these reactions are the key sites for regulation. Their activation or inactivation leads to changes in the levels of other metabolites in the pathway, including the substrates and/or products of the enzymes that catalyze "reversible" reactions. This results in passive but sensitive changes in the net flux at the reversible reactions.

This view, however, is not supported by experimental data. Genetic studies using gene dosage mutants or reverse genetics have shown that pathway flux is often unaffected by small changes in the amounts of enzymes that catalyze irreversible reactions. They have also shown that small changes in the levels of proteins that catalyze reversible reactions and nonenergized metabolite exchanges between different compartments can sometimes have a large impact on flux (Stitt and Sonnewald, 1995; Stitt, 1999; Geigenberger et al., 2004; for a recent general review, see Morandini, 2009; for an overview of photosynthetic carbon metabolism and potato [*Solanum tuberosum*] tuber starch metabolism, see Stitt et al., 2009).

The debate about the contributions of enzymes that catalyze irreversible and reversible reactions to the regulation of metabolism is also overly simplistic and confuses at least two issues.

The first issue is how metabolic pathways are structured to cope with the very high negative standard free energy ($\Delta G°'$) of some of the reactions that they contain (Stitt, 1999; Hofmeyer and Cornish-Bowden, 2000; for an extended discussion, see Morandini, 2009). With glycolysis as an example, hexokinase, ATP-phosphofructokinase, and pyruvate kinase have large negative $\Delta G°'$ values. These reactions must be held far from equilibrium in vivo, because otherwise a cell that contained a few millimolar Glc would contain several molar pyruvate, which would be incompatible with effective metabolism and probably fatal. Restricting their activity by a very low V_{max} activity (i.e. a small and limiting amount of the protein) would be incompatible with a flexible metabolic response. The activities of such enzymes are typically restricted by product, allosteric, and/or posttranslational feedback regulation. These can be relaxed, allowing their activity to rise rapidly when necessary.

The second issue is how metabolic flux is regulated. Feedback loops that act on enzymes that catalyze irreversible reactions can obviously be used, in combination with feed-forward loops, to regulate flux. However, there is no a priori reason why this is the only way to regulate metabolic flux. Indeed, exceptions to the so-called rule can be found. An example in central plant metabolism is pyrophosphate:Fru-6-P phosphotransferase. This enzyme catalyzes the reversible interconversion of Fru-6-P and Fru-1,6-bisP and is regulated by the signal metabolite Fru-2,6-bisP (Stitt and Sonnewald, 1995). Another example is in the asparate family amino acid biosynthesis pathway, where the first reaction, which is catalyzed by Asp kinase, is subject to feedback regulation in bacteria (Chassagnole et al., 2001) and Arabidopsis (Curien et al., 2009), even though this reaction is very close to equilibrium in vivo. There is also no a priori reason why enzymes that catalyze reversible reactions must be present in large "excess." As already indicated, they catalyze the forward and reverse reactions simultaneously, and an increase in net flux requires an increase of the substrate-product ratio. Such enzymes will start to restrict pathway flux if the increase of the substrate or decrease of the product impacts negatively on the operation of other enzymes in the pathway. An example in plants is the cytosolic phosphoglucoseisomerase during photosynthetic Suc synthesis. A relatively small decrease in activity leads to an increase of Fru-6-P and an amplified increase of Fru-2,6-bisP, which inhibits the cytosolic Fru-1,6-bisphosphatase and decreases the rate of Suc synthesis (Neuhaus et al., 1989).

It has often been proposed that the most effective way to alter flux is to change the activities of many or all of the enzymes in a pathway (Kacser and Acerenza, 1993; Fell and Thomas, 1995). This can occur as a result of coordinated changes in gene expression, parallel posttranslational modification of many enzymes, or recruitment of enzymes into complexes (Graham et al., 2007; Sweetlove et al., 2008). Coordinated multisite responses provide attractive paradigms to link the regulation of metabolism with the regulation of gene expression and with cellular signaling pathways (Sweetlove et al., 2008). In many cases, metabolic pathways are highly branched, in which case it is often necessary to alter fluxes through part of the network while leaving them unaltered or decreasing them in other parts of the network (Curien et al., 2009). One way in which this can occur is via "regulation by demand," in which products exert feedback regulation on the branches that produce them and shared parts of the network (Hofmeyer and Cornish-Bowden, 2000).

Application of the "Crossover" Theorem to Identify Regulated Reactions

It is necessary, therefore, to obtain experimental data about the regulatory structure of metabolic networks. Does a given response involve modulation of one or a relatively low number of enzymes, with the remainder of the network responding to the changes triggered by these enzymes, or are many enzymes modulated in parallel? Comprehensive quantitative measurements of all the metabolites in a sector of metabolism provide valuable information about the site(s) of regulation and allow hypotheses to be formulated about which regulation mechanisms are operating.

As already discussed, pathways typically contain some reactions that are close to equilibrium and others that are strongly displaced from equilibrium. For equilibrium reactions, the net flux depends on the difference between the rates of the forward reaction and the reverse reaction ($v = v^{+1} - v^{-1}$). The closer the reaction is to equilibrium, the smaller the changes of effectors (substrates, products, regulators) that are needed to produce a given change in the net flux. This obviously contributes to the flexibility of a metabolic network. While it is often technically difficult to detect the changes of metabolites that are responsible for observed changes of flux, improved precision will make this possible for an increasingly large number of reactions (Bennett et al., 2009).

For enzymes that catalyze an irreversible reaction, an increase in the in vivo activity will require either (1) an increase of the level of its substrate(s) or (2) shifts in the levels of regulatory effectors (this could include posttranslational regulation of the enzyme) that allow activity to be increased even though the substrate level decreases (e.g. a decrease in the K_m). Rolleston (1972) proposed that it is possible to pinpoint which enzymes have been activated by identifying metabolites that show reciprocal changes to pathway flux. By comparing the known properties of the enzyme with the changes of a comprehensive set of metabolites and, where appropriate, the posttranslational modification state, it is then possible to propose a qualitative model for how that enzyme is regulated. A recent example in plants is provided by Tiessen et al. (2002), who showed that the inhibition of starch synthesis after detaching potato tubers from the mother plant was accompanied by an increase of the level of the substrates of ADP-Glc pyrophosphorylase (ATP and Glc-1-P), a decrease of its product ADP-Glc, and changes of its known allosteric inhibitors that would be expected to activate the enzyme. This triggered a search for a novel posttranslational regulation mechanism that regulates its activity in response to the supply of Suc.

Analysis of Complex Metabolic Pathways Requires Modeling

With increasingly complicated metabolic networks, the application of the crossover theorem approach becomes increasingly complicated, due to the presence of branch points, cycles, and potentially redundant reactions and reaction sequences. Changes of flux are sometimes not accompanied by detectable changes of metabolites; in effect, the metabolic network operates in a self-stabilizing manner and minimizes changes in individual metabolites. Such results can point to a coordinated regulation of pathway flux acting at several sites. Three examples are the Calvin-Benson cycle, where metabolite levels are remarkably independent of the net rate of photosynthesis (Arrivault et al., 2009), the Asp family of amino acids (Curien et al., 2009), and the response of potato tuber metabolism to changes of oxygen (Geigenberger et al., 2000; Geigenberger, 2003).

Comprehensive analysis of the response of all or almost all pathway metabolic intermediates is nevertheless a powerful approach to gain insights into the regulatory structure of a metabolic network, especially when it is combined with information about enzyme kinetics to generate quantitative metabolic models. This is nicely illustrated by elegant studies of the regulation of the biosynthesis of the Asp family of amino acids (Curien et al., 2003, 2009). This pathway leads to the formation of four different amino acids (Thr, Ile, Lys, Met) and is characterized by numerous branch points and multiple feedback loops. Detailed information about the kinetic and regulatory properties of enzymes was integrated into a model that was tested by comparison with experimental studies of fluxes and metabolite levels. In this particular case, the model predicted that fluxes to Met, Lys, and Ile could be varied independently of each other and without large changes in the levels of these three amino acids, while Thr was predicted to show much larger changes in its levels. Mining of existing experimental studies confirmed this prediction and highlighted the possibility that changes of Thr might provide a useful signal that might be used to regulate metabolism at a higher level of integration.

Sensitivities of Enzymes to Their Effectors in Vivo

Quantitative measurements of metabolites can also be analyzed within the framework of flux control analysis (FCA; Kacser and Burns, 1973; Fell and Thomas, 1995). The most widely known application of FCA is to determine flux control coefficients. This is a system property that, put simply, describes the fractional change in flux that results when a single enzyme activity is changed by a small amount. Most experimental determinations of flux control coefficients use reverse genetics. As an example, decreasing the activity of cytosolic phosphoglucoisomerase by 36% leads to an 8% decrease in the rate of Suc synthesis in low light, corresponding to a flux control coefficient of 0.22 (Neuhaus et al., 1989). However, experimental determination of flux control coefficients requires a pragmatic simplification. Strictly speaking, FCA deals with infinitesimally small changes of en-

zyme activities and fluxes. In practice, quite large changes must be analyzed, due to limitations imposed by the precision and sensitivity of the analytic techniques and inherent biological variation. Curien et al. (2009) illustrate how modeling can be used to derive flux control coefficients.

The mathematical formalism of FCA also defines local properties, termed elasticity coefficients. An elasticity coefficient is the fractional change in enzyme activity that is produced by a fractional change in the level of an effector (i.e. the sensitivity of the enzyme to changes in the levels of the effector). Elasticity coefficients can be experimentally determined in vivo by introducing a small perturbation, measuring the resulting changes in flux and metabolite levels (Kacser and Burns, 1973), and introducing these into an equation that lists all known effectors of the enzyme. This does not require prior information about the detailed kinetic properties. However, each effector introduces another unknown into the equation, and a solution usually requires simplifications and decisions about which effectors are likely to be most important in vivo. An early example of the application of this approach to plant metabolism is given by Neuhaus et al. (1989, 1990), where two different approaches were used to estimate the sensitivity of the cytosolic Fru-1,6-bisphosphatase to changes in the levels of its substrate and the inhibitory signal metabolite, Fru-2,6-bisP. This is a potentially powerful approach, but it requires precise quantification of the responses of all relevant metabolites and pathway flux in enough perturbations to allow the equations to be solved in a robust manner. Recently, theoretical approaches have been developed to estimate elasticity coefficients from large perturbations, where the signal-to-noise ratio is more favorable, and to take into account side effects of the perturbations (Nikerel et al., 2006; Link and Wuester-Botz, 2007).

Large-Scale Combination of Information about Metabolite Levels with Information about the Enzyme Properties

An impressive example of how large-scale metabolite data sets can be used to generate fundamental insights into metabolism is provided by recent work of Rabinowitz and colleagues in *Escherichia coli* (Bennett et al., 2009). Using a newly developed liquid chromatography-tandem mass spectrometry (LC-MS/MS) platform (see below for more discussion), they quantified over 100 metabolites in *E. coli*. This allowed calculation of the in vivo ΔG for 25% of the reactions in *E. coli* metabolism, showing that about two-thirds occur with a loss of more than 10 kJ mol^{-1}. Another striking finding was that most metabolites in glycolysis are present at levels that are higher than the K_m values of the enzymes that metabolize them. This simplifies modeling because it implies that the key regulatory events are feedback and feed-forward loops rather than competition for substrates. This analytic platform was then used to analyze and model ammonium assimilation in *E. coli* (Yuan et al., 2009). It can be anticipated that large quantitative metabolite data sets will soon be combined with quantitative information about the concentrations of proteins and enzyme-binding sites and that this will provide important insights into the importance of enzyme kinetics as compared with metabolite diffusions, channeling, and the formation of protein complexes for the regulation of metabolic flux.

Integrating Metabolism with Growth

The biological function of metabolism is to support growth, the synthesis or turnover of storage compounds, or the accumulation of metabolites that have a role in coping with abiotic or biotic stress. In a systems approach to metabolism, therefore, it is essential to integrate the analysis of metabolism with an analysis of these higher level processes. A formal approach to this problem has been taken by Hofmeyer and Cornish-Bowden (2000), who view cellular metabolism as an economic process that is organized into supply and demand blocks. Fluxes in central metabolism can be driven by increased demand, or growth can be driven by increased supply of resources. The integration of metabolism can then be quantitatively described in terms of the elasticities/sensitivities of these two blocks of reactions. It is possible, of course, that both occur; for example, that an increased supply of carbon, nitrogen, or phosphate stimulates not only central metabolism but also, and even more strongly, protein synthesis, leading to a decrease in the level of many amino acids (Scheible et al., 2004; Morcuende et al., 2007; Osuna et al., 2007).

A complementary approach is to integrate growth processes into metabolic models. At the simplest level, the stoichiometries and fluxes of carbon, nitrogen, and ATP can be compared in central metabolism and in growth (or a particular component of growth; Piques et al., 2009). At a more sophisticated level, the fluxes to growth can be incorporated into kinetic models of metabolic pathways (Curien et al., 2009). It is also possible to analyze global metabolic responses to identify integrative or emergent characters that are associated with, or are predictive of, growth (Meyer et al., 2007; Sulpice et al., 2009; see below for more discussion). The actual costs of producing the various cellular components need to be rigorously modeled (Carlson, 2006; Barton et al., 2008). The molecular machinery that is required for growth processes can, itself, be a further major cost factor. For example, in a growing yeast cell, ribosomal proteins account for 50% of the total protein, and ribosome biogenesis may account for an even larger proportion of the total cellular energy costs (Warner, 1999). Similar levels of ribosomes may also be found in growing plant cells (Piques et al., 2009). Several different sorts of whole plant growth models are available (http://csbe.bio.ed.ac.uk/software.php) into which metabolism and

growth processes could be integrated. However, integration of models will also depend on the development of consistent and unambiguous notation (Le Novère et al., 2009).

Obtaining Quantitative Measurements of Metabolites

The previous section argued that systems biology requires comprehensive coverage and quantitative data. Paradoxically, high-throughout metabolite profiling has probably decreased the quality of metabolic data sets with respect to these two criteria, even though it has undoubtedly increased the overall scope and amount of metabolic data. Taking the Calvin-Benson cycle as an example again, the pioneering studies that led to the elucidation of the Calvin cycle in *Chlorella* employed $^{14}CO_2$ labeling followed by two-dimensional paper chromatography and autoradiography (for recent reviews, see Benson, 2002; Bassham, 2003). This platform allowed reliable detection and quantification of almost all of the metabolites in the cycle (Krause and Bassham, 1969). Studies of photosynthesis during the last 30 years have used an increasingly restricted set of enzymatic assays to detect a subset of metabolites, like glycerate-3-phosphate, dihydroxyacetone phosphate, Fru-1,6-bisP, and hexose phosphates. In the last decade, with the shift to unbiased profiling of metabolites by gas chromatography (GC)-MS or LC-MS, most data sets provide next to no information about this important sector of metabolism. These methods provide little information about phosphorylated intermediates, due to their low volatility, their low concentrations, and the difficulty to distinguish between the isomers. Arrivault et al. (2009) recently developed a platform that covered about 40 metabolites from photosynthetic metabolism, including almost all of the metabolites quantified by Calvin and coworkers (Benson, 2002; Bassham, 2003) half a century ago. To achieve this coverage, it was necessary to use two different LC systems in tandem with MS/MS and to combine this with enzymatic determinations of some metabolites that are not reliably quantified with these modern techniques.

A similar degradation of the scope and quality of metabolite data can be seen for other pathways, the exceptions being those where focused research interests have driven the time- and often material-intensive development of targeted analytic platforms. Putting it provocatively, untargeted high-throughput approaches allow the machine to dictate which metabolites are detected, while targeted approaches typically require a combination of technologies, making them time consuming and expensive. Some recent developments, however, hold great promise for extending the scope of metabolite analysis. One is Fourier transform-MS (Kaderbhai et al., 2003; Giavalisco et al., 2008; Iijima et al., 2008), where the unprecedented precision of mass detection allows a far larger number of peaks to be resolved and analyzed than in GC- or LC-MS approaches. However, the sensitivity is lower than for

dedicated LC-MS/MS approaches. Another is the development and combination of increasingly powerful and sensitive LC-MS/MS applications, exemplified by a platform that allows absolute quantification of over 100 metabolites from central metabolism in *E. coli* (Bennett et al., 2007; see below), and platforms that combine technical approaches to cover hundreds of metabolites from plant central and secondary metabolism (Akiyama et al., 2008).

Ultimately, it is unlikely that a single analytic platform will allow comprehensive coverage of all of the metabolites in all, or almost all, of the metabolic pathways in an organism. This will require the integration of data from different analytic platforms. This represents a fundamental difference between metabolomics and transcriptomics or proteomics. It will not only increase the logistical complexity for different types of metabolites but also brings issues of data integration analogous to those found when different types of data are fused (Smilde et al., 2005a; see above).

Validation of Measurements of Metabolite Levels

It is not sufficient to have analytic techniques; they must also be used correctly. In the case of metabolites, this requires showing that the reported levels of metabolites actually reflect the levels present in the biological material. There are three major sources of error. First, metabolites may change during the harvest and quench of the biological material. Second, metabolites may be lost or modified during the preparation of the extracts. Third, the compound may not have been reliably detected or quantified.

Avoiding Changes of Metabolite Levels during the Harvest and Quench

Scientists with a background in analyzing transcripts and proteins may pardonably underestimate the errors that can be introduced by inappropriate protocols for harvesting and quenching biological material when it is to be used for measurements of metabolite levels. The key issue is that metabolites often turn over much more quickly than other cellular components. If the harvest and quench takes longer than the turnover times of the metabolites, the elegant downstream analysis only reveals what is left in the stable after the horse has bolted. To take an extreme example, many Calvin-Benson cycle intermediates have turnover times of less than 1 s, and some less than 0.1 s (Stitt et al., 1980; Arrivault et al., 2009). If the light intensity is decreased for even a fraction of a second during the harvest procedure, their levels move toward those found in the dark. The quenching procedure must also be fast enough to stop metabolism in a fraction of a second. The state-of-the-art method for harvest and quenching in photosynthesis research is to freeze clamp leaves in situ in a gas-exchange chamber using electrically or pneumatically driven pistons with heads that have been precooled in

liquid nitrogen. Although turnover rates may be slower in other metabolic processes and tissues, it can take time to access tissues and cut them into thin slices, which can be quickly cooled (if the cooling process is slow you will merely learn what effect a slow cooling process has on metabolism). For example, after cutting discs from potato tubers, the increased access to oxygen leads to a 2- to 3-fold increase of the ATP/ADP ratio within a few seconds (unpublished data). Of course, many metabolites, especially those present at high levels or those involved in cellular structures, do not show such rapid changes. It is nevertheless essential to survey the literature and, if necessary, to perform control experiments to obtain estimates of the likely turnover times and design the harvest and quench procedure accordingly.

Preparation of Extracts for Analysis

After quenching, the extract must be converted into a form in which metabolites are stable and that can be applied to the analytic machine. This usually involves chemical inactivation of enzymes. In addition to specific enzymes, it is also essential to inactivate general activities like phosphatases and hydrolases that are often present at high activities in plant tissues and, thanks to their vacuolar location, are also often rather tough. The most effective method is probably still extraction and incubation at subzero temperatures by TCA in a methanol-methanol-water mix, followed by partitioning of the metabolites by repeated washing. However, this procedure is time consuming, and the traces of trichloroacetate that remain may be incompatible with many machines. Other inactivation methods, therefore, are typically used, including methanol-chloroform or methanol-alkali treatments.

Validation of the Analytic Procedure

The analytic procedure should also be validated. This requires showing that the metabolite has been correctly identified and correctly quantified. One general source of error is loss of linearity due to applying too much extract to the analytic system. This can be controlled for by establishing that the signal is strictly linear with the amount of applied extract. Other sources of error can depend on the analytic technique. Examples include incomplete extraction, incomplete derivatization (for GC-MS), and variation between machines and columns. These sources of errors need to be systematically identified and minimized (Gullberg et al., 2004). A source of error with all mass spectrometric methods is so-called "ion suppression" caused by incomplete and variable ionization of a given parent ion, due to the presence of other interfering substances. A very general and important control is to add authentic standards to the extract; in the case of MS analyses, this is best done using stable isotope-labeled standards, which

have a different mass, allowing them to be separately measured and compared with the endogenous metabolite (Fernie et al., 2004). However, there is only a limited range of commercially available stable isotope-labeled standards, and their synthesis is time consuming. Bennett et al. (2007) developed a strikingly simple solution to this problem by growing their *E. coli* on fully ^{13}C-labeled substrates, allowing readily available ^{12}C chemicals to be used as standards. A similar approach could be taken with plants, provided that gas-tight growth systems can be developed to allow plants to be grown on $^{13}CO_2$ at a reasonable price.

Recovery Experiments

Irrespective of the methods used for extraction and analysis, it is absolutely essential to validate the entire procedure by performing recovery experiments. Small amounts of authentic standards, at similar amounts to those present in the biological material, should be added to the biological material before extraction, and recovery through the extraction and analytic procedure should be documented. This is rarely done, and those studies that have been performed with plants (Roessner-Tunali et al., 2003; Schauer et al., 2006; Cruz et al., 2008; Arrivault et al., 2009; Lytovchenko et al., 2009) show that losses of metabolites can occur and that these may be avoided by modifying the extraction protocol or analytic approach.

Spatial and Temporal Heterogeneity of Metabolism and Growth

Spatial Heterogeneity

The metabolic status of a plant varies spatially and temporally. There are large differences in structural functions between tissues and the cells comprising them. Within a single cell, the presence of plastids, mitochondria, peroxisomes, and a vacuolar compartment divides the metabolites into pools that have very different properties, complicating further the interpretation of the data. At the organ level, plants can be divided between the photosynthetic tissues (source tissues), which are autotrophic, heterotrophic tissues (sink tissues), and the veins, which allow transport between the plant organs. Sink tissues include roots, the stems to some extent, parts of the young growing leaves, and the meristems, which are functionally very different. Growth is not uniformly distributed over these diverse tissues and organs. Organs growing are at least partly sink tissues, their development depending on the availability of resources provided by source tissues (Turgeon and Wolf, 2009). A large proportion of the plant does not contribute to growth after establishment, and those parts that do are regulated by environmental cues such as seasonal and diurnal rhythms and stresses, which affect the availability for resources.

Temporal Heterogeneity: The Impact of Differing Dynamic Responses on the Momentary Metabolite Profile

Photosynthetic tissues are autotrophic only during the light period. They store reduced carbon for the night period, when there is a requirement for maintenance but also for growth (Schurr et al., 2006; Walter et al., 2009). This is largely performed by accumulating a transient pool of starch in the light and remobilizing it during the night, this process being highly regulated (Smith and Stitt, 2007). Growth is also not linear in time or uniformly distributed over the diverse tissues and organs. This regulation is in part diurnally regulated.

However, plants are also exposed to a multitude of short-term environmental fluctuations. They are also exposed to slower changes, which occur over weeks or on a seasonal basis. These include daily differences in climatic conditions such as light, temperature, and water availability and biotic interactions. While the clock may allow the response to diurnal and seasonal shifts to anticipate changes before they actually take place, the response to short-term fluctuations cannot be anticipated but requires instantaneous regulatory responses.

As already mentioned, many metabolites are very susceptible to short-time environmental changes, which is not, or much less, the case for transcripts and proteins (Gibon et al., 2004a). Many aspects of metabolism also have a very different dynamic temporal range to that of growth. The same holds for the relation between metabolites that are intermediates in assimilatory and biosynthetic pathways and metabolites that are part of cellular structures. This implies that (1) the existing available metabolic machinery determines largely the adjustments of the metabolite status to the changing environments and (2) is able to perceive and respond to environmental signals (Gibon et al., 2004b; Piques et al., 2009). Growth shows a range of responses to short-term perturbations. While root growth responds within minutes to environmental perturbations (Nagel et al., 2006), shoot growth may not respond so quickly (Walter et al., 2009). Possibly, the highly fluctuating conditions met aboveground mean that rapid changes in shoot growth are detrimental, due to the energy cost of making and degrading the growth machinery. Taken together, these results suggest that a simple link between metabolic status and growth can be expected only when plants are facing stable conditions. More generally, the high temporal and spatial heterogeneity for metabolites and growth, as well as the importance of the regulation of the transport of resources between source and sink tissues, must be taken into account when designing experiments to assess potential links between the metabolic activity, metabolite composition, and growth.

How to Perturb the System?

To generate a metabolic network, a perturbation must be used to generate variation in the metabolite contents. This can be done by applying the perturbation at the environmental, chemical, and/or genetic level.

Environmental Perturbations: Steady States Versus Dynamic Dimension

Metabolic networks can be generated by comparing metabolic profiles obtained for different stress intensities from plants treated for the same duration and harvested at the same time (steady states) or by following the metabolic response to a changed constraint over the time (dynamic range).

An interesting example of the use of steady-state conditions is a study of Arabidopsis ecotype Columbia submitted to either short- or long-term nitrogen, sulfur, or sulfur and nitrogen deficiency (Hirai et al., 2004). The authors combined metabolic and transcriptome data and identified common as well as specific responses for the different nutrient stresses. However, as the experimental design was restricted to steady-state levels of transcriptome and metabolome, it provided only limited temporal information, which is often needed to draw conclusions about how processes are regulated. Another general drawback is that it is often difficult to build correlation networks, due to a large excess of traits over treatments and to the nonlinearity of the changes between the components of the metabolome.

Time-series experiments reveal temporal dependencies between metabolic traits. They also provide information about behavior during the time in which a new steady state is established, which can provide insights into functional dependencies that are masked in steady states. Some highly regulated metabolites may be maintained at near homeostasis in different conditions; for such metabolites, the variation between different steady states is likely to be too small to allow dependencies to be perceived. Time-course experiments have been used to gain insights into relationships between enzyme activities, metabolites, and transcripts (Gibon et al., 2006). The diurnal variation of these three components was measured in wild-type Arabidopsis and in the starchless *phosphoglucomutase* (*pgm*) mutant, which accumulates large amounts of sugars in the light but experiences carbon starvation in the night. The transcript levels showed large diurnal variation, which were further enhanced in the mutant compared with wild-type plants. Enzyme activities and metabolites showed smaller diurnal changes and were in a similar range both in the wild type and *pgm* (except for the sugars). This surprising result suggests that correlations between metabolites and transcripts, which are often interpreted in the literature as the result of a modification of metabolism resulting from changes in gene expression, may often be due to the regulation of transcript levels by metabolic status.

Until now, time-course studies in plants do not have enough time data points to build a network. They serve to identify general trends for adaptation to

environmental conditions and are subsequently extended by characterization of the potential candidate genes/pathways using mutant/transgenic approaches. In view of the vast number of genes probably involved, it is time consuming to obtain a comprehensive overview. Moreover, the concept of the "emergent properties" of networks assumes that the properties of a network are not necessarily equal to the sum of the properties of the traits involved (Corning, 2002). Thus, functional characterization of a gene by a transgenic/mutant approach provides information about its potential function(s), but these may differ from the functions when the gene is embedded in a different network. This might explain why genes identified by network analyses often appear to have no or very little effect when they are studied by genetic approaches.

Chemical Genomics

Chemical genomics provides a powerful tool to identify protein-phenotype relationships. Chemicals can be used in a targeted way when the site of action of the chemical is known or in an untargeted way by screening large chemical libraries for a desired phenotype. Targeted approaches have been widely used for many years. This is a powerful approach, provided care is taken with respect to the specificity of the chemical, the concentration range, and the mode of application. For example, after showing that geldanamycin can be used to inhibit heat shock protein 90, Queitsch et al. (2002) and Sangster et al. (2008) applied this chemical to an Arabidopsis RIL population and natural accessions and uncovered large genetic variation for root and hypocotyl growth that is presumably masked by the action of this heat shock protein. Chemicals can also be used in a reverse untargeted way. There, the mode of action of the chemical is not postulated per se, and large chemical libraries can be screened for the desired phenotype. This approach has of course been used for long time to search for herbicides, whose mode of action is subsequently elucidated. In an elegant recent application, Cutler and colleagues screened for synthetic seed germination inhibitors that differentially affect Arabidopsis accessions (Zhao et al., 2007) and used one of the discriminating molecules (pyrabactin) to identify ABA sensors (Park et al., 2009).

Reverse Genetics

One classic intervention is to decrease the expression of individual enzymes in a pathway. The most common approach is to use knockout mutants or techniques (e.g. RNA interference, artificial microRNA) that produce a large decrease in the activity of an enzyme. This is a powerful and valuable method to identify essential genes. It has been used extensively by Saito and colleagues to identify novel gene functions, including the characterization of different genes of the Ser acetyltransferase family (Watanabe et al.,

2008) and a chloroplastic UGPase involved in the first step of sulfolipid synthesis (Okazaki et al., 2009), and the involvement of some clock-associated pseudoresponse regulator proteins in the circadian regulation of the mitochondrial metabolism (Fukushima et al., 2009).

A complementary approach is to increase the activity of the gene of interest. However, many genes are expressed differentially or even specifically in some tissues/cells, and expressing them in tissues where they are normally not expressed can lead to pleiotropic effects. This can be avoided by the use of tissue-specific promoters, which are beginning to be available for a large set of cells/tissues. If the genes of interest are potentially involved in sensing/signaling, it might be better to study their effect in a dynamic range using inducible expression. The latter can also be very useful when constitutive expression leads to developmental defects and especially difficulties in obtaining viable seeds.

Genetic interventions that result in large changes are not a method of choice to analyze metabolic networks, because the response is likely to involve a major change in the network (Corning, 2002; Last et al., 2007; Sweetlove et al., 2008). An alternative approach is to use genetic strategies that result in a partial decrease in target gene expression. The ideal genotypes for analysis of metabolic networks are often those in which fluxes are unchanged or only slightly changed. The first technology for decreasing gene expression in plants, so-called "antisense" constructs, typically resulted in a set of transgenics with a progressive decrease in expression of the targeted gene (Stitt and Sonnewald, 1995). Retrospectively, this technology was appropriate for enzymes that are encoded by single genes or small families with no differentiation of function, like many of the Calvin cycle enzymes. It was less appropriate when the members of a gene family produce isoforms with differing kinetic or regulatory properties or expression patterns.

A generic and gene-specific strategy would be to quantitatively analyze responses in heterozygotes of knockout mutants. Quantitative analysis of the response of transcript, protein, and metabolite levels and flux to a loss of one of the two functional genes in a diploid organism would provide a data set that defines the contribution of transcriptional regulation, regulation of protein turnover, and metabolic regulation to the regulation of a given enzyme and allow integrated analysis across these functional levels. As pointed out by Kacser and Burns (1973), transcript and protein levels are indeed typically halved in heterozygotes with one functional and one nonfunctional gene copy. However, there are notable exceptions, and these identify enzymes where changes of expression, translation, and protein degradation are tightly integrated with metabolic regulation. For example, tobacco (*Nicotiana tabacum*) mutants with a decreased number of functional gene copies for nitrate reductase have an increased number of *NIA* transcripts (these include

transcripts from the nonfunctional genes, which are still transcribed), modified posttranslational regulation of the NIA protein, and altered dynamics of the NIA protein during diurnal cycles (Scheible et al., 1997), which are accompanied by modified levels and temporal dynamics of the transcripts and activities of several other enzymes of central nitrogen and carbon metabolism (Scheible et al., 2000). However, this approach places a high premium on the experimental design and the quantitative precision of the analytic platforms.

Natural Diversity

Many of the genetic perturbations discussed above also occur in a more natural setting. Natural diversity provides a large number of changes, presumably often of functional importance and, when many different genotypes coexist, in a large number of different combinations (Jansen, 2003b). Species-wide information about polymorphisms (Clark et al., 2007; Weigel and Mott, 2009) will soon become available as a result of genome resequencing for Arabidopsis, and similar information resources will follow for several major crops. This will place a premium on the generation of data matrices about metabolites and other physiological traits, which can be combined with comprehensive in silico information. This approach will become an important interface between basic research and crop breeding.

Sequence variations within species range from single nucleotide polymorphisms to insertions/deletions and even chromosomal rearrangements like duplication and translocation. The effect of such alterations can vary from neutral to structural diversification of protein function but also to differences in transcription and response to environmental cues. Duplication events, for instance, are a major source of neofunctionalization and subfunctionalization in gene families (Bikard et al., 2009; Kliebenstein, 2009), whereas point mutations can lead to subtle quantitative differences in biological processes (Fridman et al., 2004). Mutations that have accumulated during the evolutionary history of a species might persist in different lineages or radiate in outcrossing species through genomic recombination and selection. The pool of heritable structural DNA variation within a species is commonly referred to as natural diversity and is thought to play an important role in diversification and adaptation to a wide variety of growth habitats (Mitchell-Olds and Schmitt, 2006; Alonso-Blanco et al., 2009).

Because of the large variation in phenotypic traits, natural diversity is often used in genetic studies that aim to unravel the mechanistic basis of physiological processes. This has been especially fruitful for quantitative traits in which phenotypes are the result of multiple, possibly interacting, genes. Polygenic regulation often results in a continuous distribution in trait values, making it difficult to associate phenotypes with genotypic variants, in contrast to qualitative traits, where phenotypic values can be assigned to a limited number of discrete classes. To this end, so-

phisticated statistical and experimental approaches have been developed that are generally referred to as quantitative trait locus (QTL) analysis (Broman, 2001; Doerge, 2002; Slate, 2005). In QTL analysis, a population of plants, in which genetic variation for the trait of interest is present, is genotyped with evenly distributed molecular markers, after which linkage is sought between phenotypic values and the genotype at the marker positions. A strong association between the phenotype and genotype indicates genomic loci responsible for the variation observed. Several types of populations are used for QTL analyses, with the largest distinction made between experimental and natural populations.

Generation and Use of Experimental Populations

Experimental populations are the result of the intercrossing of two or more distinct genotypes. The exact structure of the population after generating the initial F1 depends on the applied crossing scheme, which can be repetitive selfing, backcrossing, or advanced intercrossing. Each method results in a different genetic makeup of the population, each with its own advantages and disadvantages (Jansen, 2003a). RILs are generally constructed by crossing two distinct genotypes and repetitive selfing of the progeny into a population of (nearly) homozygous individuals. Each individual then consists of a different mosaic of the two genotypes, and multiple QTLs segregate independently in the population. This allows for the simultaneous detection of each contributing QTL and an estimation of epistatic effects of multiple loci. Many quantitative traits, however, are determined by a large number of small-effect QTLs, which might be masked by the noise introduced by large-effect loci. Therefore, near-isogenic lines (NILs) or introgression lines are also frequently used (Eshed and Zamir, 1995; Monforte and Tanksley, 2000). NILs only contain a single introgression of a donor genotype in an otherwise homogenic background and are generated by repetitive backcrossing of the F1 with the recurrent parent. This way, QTLs located in introgression regions are Mendelized and their detection is not impaired by the independent segregation of additional QTLs, but no information can be gained of the interaction between QTLs. Both types of populations consist of homozygous lines that make them immortal and enable the use of identical genotypes in replication and different experimental settings. Immortal populations, however, provide no information about additive and dominant QTL effects, a reason that some populations are not selfed to full homozygosity (e.g. F2; Loudet et al., 2005; Semel et al., 2006; Schauer et al., 2008).

Although experimental populations have been extensively used in forward genetics studies, surprisingly little work has been done on the role of primary metabolism in growth and development. This is partly because growth is a highly complex trait and the final

outcome of the interplay of many physiological processes and their interaction with the environment. Several studies have reported on the relationship between growth and developmental timing, plant morphology, nutrient use efficiency, and robustness toward environmental perturbations. Only recently, however, have attempts been made to link differences in growth to variation in primary metabolism (Causse et al., 2004; Calenge et al., 2006; Keurentjes et al., 2006, 2008; Schauer et al., 2006; Meyer et al., 2007; Lisec et al., 2008; Sulpice et al., 2009). These studies confirmed that growth is not regulated by a single factor but is accompanied by the simultaneous regulation of a large part of primary metabolism, suggesting genetic control of a dense and robust biochemical network.

The use of natural diversity in experimental populations can be particularly valuable here, because it provides subtle changes in subparts of the network and allows the identification of genetic factors explaining the observed variation. As such, it can identify directionality in relationships (Szymanski et al., 2007; Jansen et al., 2009; Kliebenstein, 2009), a property that is often missed in correlation studies where no information about cause and consequence can be obtained. Immortal populations also allow destructive analyses of different tissues, developmental stages, and experimental settings, revealing information on spatial, temporal, and conditional regulation.

A major drawback of artificial populations, however, is the segregation of only a limited number of allelic variants and the low mapping resolution due to the restricted number of recombination events. The latter can only be overcome by advanced intercrossing and severely increasing the population size (Zou et al., 2005; Balasubramanian et al., 2009). Furthermore, a comment often made on experimental populations concerns the artificial nature of recombination of genomes. Since such genotypes are neither created nor selected for by nature, this imposes the question of how relevant the findings are in an ecological and evolutionary context.

Another limitation is that when a large numbers of traits are measured (e.g. by metabolite profiling), QTLs often colocalize to the same genomic segment. It is tempting to interpret colocalizations as causal (i.e. due to the same genetic change) and to use them to build up a metabolic network. However, colocations may be fortuitous, due to the large size of the genomic sectors demarcated in RIL and NIL lines and the high frequency of polymorphisms in plant genomes. Further analysis requires fine-mapping and identification of the specific gene and the causal genetic polymorphism. This is laborious. While it can be done for single genes (Fridman et al., 2004), it is a major undertaking for tens or hundreds of QTLs.

Association Mapping in Natural Populations

Genome-wide association (GWA) or linkage disequilibrium (LD) mapping make use of the much larger number of alleles and historical recombination events in natural populations. Because of the high recombination frequency in natural populations, an equally higher number of genetic markers are needed to detect LD, but mapping resolution is much higher when such LD is detected. For a number of species, LD is estimated to be close to the actual gene density, allowing direct identification of the genes underpinning detected QTLs (Nordborg et al., 2002; Kim et al., 2007). Through the technological advances made in marker technology, GWA now becomes within reach in plant sciences (Harjes et al., 2008; Yu et al., 2008), although as yet no studies have reported on its use for growth and metabolism (Nordborg and Weigel, 2008). Nonetheless, promising results have been obtained with local association mapping, in which a single gene or short genomic region is sequenced for a large number of varieties (Sterken et al., 2009). Thus, classes of sequence variants, or haplotypes, could be assigned to variation in growth and related to primary metabolism (Wilson et al., 2004; Sulpice et al., 2009). In some cases, the causal nucleotide polymorphism could even be identified. This may require additional layers of biological information to short list the candidate genes (Sulpice et al., 2009). These pioneering studies indicate that much progress can be made in the genetic dissection of the regulation of growth.

However, there are a number of caveats to be considered for GWA studies. A first remark concerns the bias in findings when using natural diversity in genetic studies, which applies to both artificial and natural populations. QTLs can only be detected when they display variation in a population, meaning that highly conserved genetic factors, which are probably the most essential, will not be picked up in linkage studies. A second concern when using GWA in natural populations is the presence of population structure. Population structure is the result of the relative genetic isolation of subpopulations in which variation in traits accumulates simultaneously with random noncausal mutations (Nordborg et al., 2005). This relationship is insufficiently broken due to limited recombination with geographically distant subpopulations and can result in many false-positive associations. This can be partly overcome by applying structure analyses (Bradbury et al., 2007), but when variation in a trait is not uniformly distributed over the mapping population, corrections for structure will also remove true positives (Pritchard et al., 2000). Structure can also be avoided by restricting GWA to confined populations (e.g. sampled from small regions or breeding pedigrees), although less natural variation will be present in such populations. Another hallmark of natural populations is the presence of selective sweeps, which occur when a beneficial mutation becomes rapidly fixed in a species (Borevitz et al., 2007; Tian et al., 2009). Because of the limited number of recombinations during fixation, large stretches of flanking sequence can also become fixed due to linkage drag. This phenomenon is known as genetic hitchhiking (Barton,

2000) and can extend LD substantially, with dramatic effects on resolution. Finally, the frequency of allelic variants and their additive effect needs to be sufficient to be able to statistically detect QTLs. In natural populations, numerous alleles of a single locus may coexist, and their effect on the trait of interest might also vary accordingly. When a unique variant is present at very low frequency, it might go undetected, especially when a high number of loci determine the eventual phenotype. Likewise, small-effect loci might not be able to be discriminated from the noise introduced by independently segregating larger effect loci. Most of these problems can partly be overcome by analyzing large population sizes and accurate estimates of trait values, but for a number of traits, GWA might not be the most suitable approach to follow and a combination of genetic approaches might be complementary (Yu et al., 2008).

It remains to be seen whether such a complex trait as growth fits the prerequisites for successful GWA mapping. The importance of growth for reproductive success is undisputed, and it is not unlikely that beneficial mutations have spread rapidly in natural populations, possibly creating selective sweeps and reducing variation. On the other hand, many tradeoffs between growth and other properties (e.g. disease resistance) of plants resulting in a balanced selection of variation have been reported (Tian et al., 2002; Kroymann and Mitchell-Olds, 2005; Chan et al., 2009). Balancing selection ensures the persistence of multiple alleles and, depending on the sampled population, in sufficient frequencies. Such properties are in favor of GWA analyses, but growth is regulated by many factors that together form a robust metabolic network. It is probable that variation in subparts of this network is buffered by other parts (Fu et al., 2009), for example, by adjusting flux rates. Such a scenario might make it difficult to relate growth to the effect of isolated QTLs without taking into account the interaction with other factors. Perhaps growth first needs to be disentangled from the various contributing components, which then can be analyzed for variation and its effect on other components. This would plead for a bottom-up approach that will analyze the effect of natural variation on the genetic regulation of basic elements like the transcription of genes, the activity of enzymes, and the accumulation of metabolites (Keurentjes et al., 2008). This should be followed by an analysis of how such alterations affect pathways, fluxes, and the allocation of resources and finally how these processes interact, culminating in the control of growth and development. Such an approach would also teach us much about the genetic mechanisms of regulation in complex traits (Mitchell-Olds and Pedersen, 1998).

CONCLUSION

Thus, a wide variety of information is generated by comprehensive analyses of metabolites. The analytic platforms to do this are becoming increasingly powerful, as the first generation of metabolite profiling with GC-MS and LC-MS is complemented by Fourier transform-MS and by platforms that combine different LC-MS/MS approaches. The evaluation of metabolite data is already embedded in a mathematical environment. This interaction will become even more important as data are integrated from different platforms. In parallel with the use of increasingly powerful technologies, it is vital to employ suitable protocols for harvesting, quenching, sample handling, and analysis and to perform control experiments in order to validate that the measurements faithfully reflect the levels of metabolites that were present in the tissue. Many of the most important uses of metabolite data require quantitative data. This will become even more so as systems biology expands and kinetic and other forms of modeling play an increasingly important role in the analysis and integration of experimental data. Modeling also depends crucially on experimental design. The issues that need to be considered include statistical design, problems and complications due to heterogeneities of metabolism in space and time, and the specificity of the interventions. However, probably one of the most powerful approaches to understand the diversity of plant metabolism and its relationship to growth will be to employ metabolite profiling in combination with the increasingly sophisticated methodologies of quantitative genetics to unravel responses of metabolism in a multivariate experimental context. This will provide a fertile interface to plant breeding and crop improvement.

LITERATURE CITED

Akiyama K, Chikayama E, Yuasa H, Shimada Y, Tohge T, Shinozaki K, Hirai MY, Sakurai T, Kikuchi J, Saito K (2008) Prime: a Web site that assembles tools for metabolomics and transcriptomics. In Silico Biol 8: 339–345

Alonso-Blanco C, Aarts MG, Bentsink L, Keurentjes JJB, Reymond M, Vreugdenhil D, Koornneef M (2009) What has natural variation taught us about plant development, physiology, and adaptation? Plant Cell 21: 1877–1896

Antoniewicz MR, Kelleher JK, Stephanopoulos G (2007) Accurate assessment of amino acid mass isotopomer distributions for metabolic flux analysis. Anal Chem 79: 7554–7559

Arrivault S, Guenther M, Ivakov A, Feil R, Vosloh D, van Dongen JT, Sulpice R, Stitt M (2009) Use of reverse-phase liquid chromatography, linked to tandem mass spectrometry, to profile the Calvin cycle and other metabolic intermediates in Arabidopsis rosettes at different carbon dioxide concentrations. Plant J 59: 824–839

Balasubramanian S, Schwartz C, Singh A, Warthmann N, Kim MC, Maloof JN, Loudet O, Trainer GT, Dabi T, Borevitz JO, et al (2009) QTL mapping in new Arabidopsis thaliana advanced intercross-recombinant inbred lines. PLoS One 4: e4318

Barton DB, Pappl B, Delneri D, Oliver SB, Rattrav M, Bergman CM (2008) Systems biology of energetic and atomic costs in the yeast transcriptome, proteome, and metabolome. Nature Precedings hdl:10101/npre.2008.1841.2

Barton NH (2000) Genetic hitchhiking. Philos Trans R Soc B 355: 1553–1562

Bassham JA (2003) Mapping the carbon reduction cycle: a personal retrospective. Photosynth Res 76: 37–52

Bassham JA, Krause GH (1969) Free energy changes and metabolic regulation in steady-state photosynthetic carbon reduction. Biochim Biophys Acta **189:** 207–221

Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. Nat Chem Biol **5:** 593–599

Bennett BD, Yuan J, Kimball EH, Rabinowitz JD (2007) Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. Nat Protoc **3:** 1299–1311

Benson AA (2002) Following the path of carbon in photosynthesis: a personal story. Photosynth Res **73:** 31–49

Bikard D, Patel D, Le Mette C, Giorgi V, Camilleri C, Bennett MJ, Loudet O (2009) Divergent evolution of duplicate genes leads to genetic incompatibilities within A. thaliana. Science **323:** 623–626

Borevitz JO, Hazen SP, Michael TP, Morris GP, Baxter IR, Hu TT, Chen H, Werner JD, Nordborg M, Salt DE, et al (2007) Genome-wide patterns of single-feature polymorphism in *Arabidopsis thaliana*. Proc Natl Acad Sci USA **104:** 12057–12062

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) Tassel: software for association mapping of complex traits in diverse samples. Bioinformatics **23:** 2633–2635

Broman KW (2001) Review of statistical methods for QTL mapping in experimental crosses. Lab Anim (NY) **30:** 44–52

Bylesjö M, Eriksson D, Kusano M, Moritz T, Trygg J (2008) Data integration in plant biology: the O2PLS method for combined modeling of transcript and metabolite data. Plant J **52:** 1181–1191

Bylesjö M, Nilsson R, Srivastava V, Grönlund A, Johansson AI, Jansson S, Karlsson J, Moritz T, Wingsle G, Trygg J (2009) Integrated analysis of transcript, protein and metabolite data to study lignin biosynthesis in hybrid aspen. J Proteome Res **8:** 199–210

Calenge F, Saliba-Colombani V, Mahieu S, Loudet O, Daniel-Vedele F, Krapp A (2006) Natural variation for carbohydrate content in Arabidopsis: interaction with complex traits dissected by quantitative genetics. Plant Physiol **141:** 1630–1643

Carlson RP (2006) Metabolic systems cost-benefit analysis for interpreting network structure and regulation. Bioinformatics **23:** 1258–1264

Causse M, Duffe P, Gomez MC, Buret M, Damidaux R, Zamir D, Gur A, Chevalier C, Lemaire-Chamley M, Rothan C (2004) A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. J Exp Bot **55:** 1671–1685

Chan EK, Rowe HC, Kliebenstein DJ (2009) Understanding the evolution of defense metabolites in Arabidopsis thaliana using genome-wide association mapping. Genetics (in press)

Chassagnole C, Rais B, Quentin E, Fell DA, Mazat JP (2001) An integrated study of threonine-pathway enzyme kinetics in Escherichia coli. Biochem J **356:** 415–423

Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA, et al (2007) Common sequence polymorphisms shaping genetic diversity in Arabidopsis thaliana. Science **317:** 338–342

Corning PA (2002) The re-emergence of "emergence": a venerable concept in search of a theory. Complexity **7:** 18–30

Cruz JA, Emery C, Wust M, Kramer DM, Lange BM (2008) Metabolite profiling of Calvin cycle intermediates by HPLC-MS using mixed-mode stationary phases. Plant J **55:** 1047–1060

Curien G, Bastien O, Robert-Genthon M, Cornish-Bowden A, Cádenas ML, Dumas R (2009) Understanding the regulation of aspartate metabolism using a model based on measured kinetic parameters. Mol Syst Biol **5:** 271

Curien G, Ravanel S, Dumas R (2003) A kinetic model of the branchpoint between the methionine and threonine biosynthesis pathways in Arabidopsis thaliana. Eur J Biochem **270:** 4615–4627

Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. Nat Rev Genet **3:** 43–52

Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics **141:** 1147–1162

Fell DA, Thomas S (1995) Physiological control of metabolic flux: the requirement for multisite modulation. Biochem J **311:** 35–39

Fernie AR, Geigenberger P, Stitt M (2005) Flux, an important, but neglected, component of functional genomics. Curr Opin Plant Biol **8:** 174–182

Fernie AR, Trethewey RN, Krotzky AJ, Willmitzer L (2004) Metabolite profiling: from diagnostics to systems biology. Nat Rev Mol Cell Biol **5:** 763–769

Fridman E, Carrari F, Liu YS, Fernie AR, Zamir D (2004) Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science **305:** 1786–1789

Fu J, Keurentjes JJB, Bouwmeester H, America T, Verstappen FW, Ward JL, Beale MH, de Vos RC, Dijkstra M, Scheltema RA, et al (2009) System-wide molecular evidence for phenotypic buffering in Arabidopsis. Nat Genet **41:** 166–167

Fukushima A, Kusano M, Nakamichi N, Kobayashi M, Hayashi N, Sakakibara H, Mizuno T, Saito K (2009) Impact of clock-associated Arabidopsis pseudo-response regulators in metabolic coordination. Proc Natl Acad Sci USA **106:** 7251–7256

Geigenberger P (2003) Response of plant metabolism to too little oxygen. Curr Opin Plant Biol **6:** 247–256

Geigenberger P, Fernie AR, Gibon Y, Christ M, Stitt M (2000) Metabolic activity decreases as an adaptive response to low internal oxygen in growing potato tubers. Biol Chem **381:** 723–740

Geigenberger P, Stitt M, Fernie AR (2004) Metabolic control analysis and regulation of the conversion of sucrose to starch in growing potato tubers. Plant Cell Environ **27:** 655–673

Giavalisco P, Hummel J, Lisec J, Inostroza AC, Catchpole G, Willmitzer L (2008) High-resolution direct infusion-based mass spectrometry in combination with whole C-13 metabolome isotope labeling allows unambiguous assignment of chemical sum formulas. Anal Chem **80:** 9417–9425

Gibon Y, Blaesing OE, Hannemann J, Carillo P, Hohne M, Hendriks JHM, Palacios N, Cross J, Selbig J, Stitt M (2004a) A robot-based platform to measure multiple enzyme activities in *Arabidopsis* using a set of cycling assays: comparison of changes of enzyme activities and transcript levels during diurnal cycles and in prolonged darkness. Plant Cell **16:** 3304–3325

Gibon Y, Blaesing OE, Palacios-Rojas N, Pankovic D, Hendriks JH, Fisahn J, Hohne M, Gunther M, Stitt M (2004b) Adjustment of diurnal starch turnover to short days: depletion of sugar during the night leads to a temporary inhibition of carbohydrate utilization, accumulation of sugars and post-translational activation of ADP-glucose pyrophosphorylase in the following light period. Plant J **39:** 847–862

Gibon Y, Usadel B, Blaesing OE, Kamlage B, Hoehne M, Trethewey R, Stitt M (2006) Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in Arabidopsis rosettes. Genome Biol **7:** R76

Goodacre R, Broadhurst D, Smilde AK, Kristal BS, Baker JD, Beger R, Bessant C, Connor S, Capuani G, Craig A, et al (2007) Proposed minimum reporting standards for data analysis in metabolomics. Metabolomics **3:** 231–241

Grafahrend-Belau E, Schreiber F, Koschutzki D, Junker BH (2009) Flux balance analysis of barley seeds: a computational approach to study systemic properties of central metabolism. Plant Physiol **149:** 585–598

Graham JWA, Williams TCR, Morgan M, Fernie AR, Ratcliffe RG, Sweetlove LJ (2007) Glycolytic enzymes associate dynamically with mitochondria in response to respiratory demand and support substrate channeling. Plant Cell **19:** 3723–3738

Gullberg J, Jonsson P, Nordstrom A, Sjostrom M, Moritz T (2004) Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of Arabidopsis thaliana samples in metabolomic studies with gas chromatography/mass spectrometry. Anal Biochem **331:** 283–295

Guy C, Kopka J, Moritz T (2008) Plant metabolomics coming of age. Physiol Plant **132:** 113–116

Hall RD (2006) Plant metabolomics: from holistic hope, to hype, to hot topic. New Phytol **169:** 453–468

Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, et al (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science **319:** 330–333

Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, Kanaya S, Nakamura Y, Kitayama M, Suzuki H, et al (2005) Elucidation of gene-to-gene and metabolite-to-gene networks in Arabidopsis by integration of metabolomics and transcriptomics. J Biol Chem **280:** 25590–25595

Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M, Fujiwara T, Saito K (2004) Integration of transcriptomics and

metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. Proc Natl Acad Sci USA **101:** 10205–10210

Hofmeyer JHS, Cornish-Bowden A (2000) Regulating the cellular economy of supply and demand. FEBS Lett **476:** 47–51

Iijima Y, Nakamura Y, Ogata Y, Tanaka K, Sakurai N, Suda K, Suzuki T, Suzuki H, Okazaki K, Kitayama M, et al (2008) Metabolite annotations based on the integration of mass spectral information. Plant J **54:** 949–962

Janes KA, Yaffe MB (2006) Data-driven modelling of signal-transduction networks. Nat Rev Mol Cell Biol **7:** 820–828

Jansen RC (2003a) Quantitative trait loci in inbred lines. *In* DJ Balding, M Bishop, C Cannings, eds, Handbook of Statistical Genetics. John Wiley & Sons, Chichester, UK, pp 445–476

Jansen RC (2003b) Studying complex biological systems using multifactorial perturbation. Nat Rev Genet **4:** 145–151

Jansen RC, Tesson BM, Fu J, Yang Y, McIntyre LM (2009) Defining gene and QTL networks. Curr Opin Plant Biol **12:** 241–246

Kacser H, Acerenza L (1993) A universal method for achieving increases in metabolite production. Eur J Biochem **216:** 361–367

Kacser H, Burns JA (1973) The control of flux. *In* Rate Control of Biological Processes. Cambridge University Press, Cambridge, UK, pp 65–104

Kaderbhai NN, Broadhurst DI, Ellis DI, Goodacre R, Kell DB (2003) Functional genomics via metabolic footprinting: monitoring metabolite secretion by Escherichia coli tryptophan metabolism mutants using FT-IR and direct injection electrospray mass spectrometry. Comp Funct Genomics **4:** 376–391

Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ, Feldblyum T, Nierman W, Benito MI, Lin XY, et al (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature **408:** 796–815

Keurentjes JJB, Fu J, de Vos CH, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D, Koornneef M (2006) The genetics of plant metabolism. Nat Genet **38:** 842–849

Keurentjes JJB, Sulpice R, Gibon Y, Fu J, Koornneef M, Stitt M, Vreugdenhil D (2008) Integrative analyses of genetic variation in enzyme activities of primary carbohydrate metabolism reveal distinct modes of regulation in Arabidopsis thaliana. Genome Biol **9:** R129

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet **39:** 1151–1155

Kind T, Scholz M, Fiehn O (2009) How large is the metabolome? A critical analysis of data exchange practices in chemistry. PLoS One **4:** e5440

Kliebenstein D (2009) Advancing genetic theory and application by metabolic quantitative trait loci analysis. Plant Cell **21:** 1637–1646

Kopka J, Fernie A, Weckwerth W, Gibon Y, Stitt M (2004) Metabolite profiling in plant biology: platforms and destinations. Genome Biol **5:** 109

Krause GH, Bassham JA (1969) Induction of respiratory metabolism in illuminated Chlorella pyrenoidosa and isolated spinach chloroplasts by addition of vitamin k5. Biochim Biophys Acta **172:** 553–565

Kroymann J, Mitchell-Olds T (2005) Epistasis and balanced polymorphism influencing complex trait variation. Nature **435:** 95–98

Kruger NJ, Ratcliffe RG (2009) Insights into plant metabolic networks from steady-state metabolic flux analysis. Biochimie **91:** 697–702

Kusano M, Fukushima A, Arita M, Jonsson P, Moritz T, Kobayashi M, Hayashi N, Tohge T, Saito K (2007) Unbiased characterization of genotype-dependent metabolic regulations by metabolomic approach in Arabidopsis thaliana. BMC Syst Biol **1:** 53

Last RL, Jones AD, Shachar-Hill Y (2007) Towards the plant metabolome and beyond. Nat Rev Mol Cell Biol **8:** 167–174

Lehmann M, Schwarzlander M, Obata T, Sirikantaramas S, Burow M, Olsen CE, Tohge T, Fricker MD, Moller BL, Fernie AR, et al (2009) The metabolic response of Arabidopsis roots to oxidative stress is distinct from that of heterotrophic cells in culture and highlights a complex relationship between the levels of transcripts, metabolites, and flux. Mol Plant **2:** 390–406

Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM, et al (2009) The systems biology graphical notation. Nat Biotechnol **27:** 735–741

Liao JC, Boscolo R, Yang YL, Tran LM, Sabatti C, Roychowdhury VP (2003) Network component analysis: reconstruction of regulatory signals in biological systems. Proc Natl Acad Sci USA **100:** 15522–15527

Libourel IGL, Shachar-Hill Y (2008) Metabolic flux analysis in plants: from intelligent design to rational engineering. Annu Rev Plant Biol **59:** 625–650

Link H, Wuester-Botz D (2007) Steady-state analysis of metabolic pathways: comparing the double modulation method and the lin-log approach. Metab Eng **9:** 433–441

Lisec J, Meyer RC, Steinfath M, Redestig H, Becher M, Witucka-Wall H, Fiehn O, Torjek O, Selbig J, Altmann T, et al (2008) Identification of metabolic and biomass QTL in Arabidopsis thaliana in a parallel analysis of RIL and IL populations. Plant J **53:** 960–972

Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. Nat Protoc **1:** 387–396

Loudet O, Gaudon V, Trubuil A, Daniel-Vedele F (2005) Quantitative trait loci controlling root growth and architecture in *Arabidopsis thaliana* confirmed by heterogeneous inbred family. Theor Appl Genet **110:** 742–753

Lunn JE (2007) Gene families and evolution of trehalose metabolism in plants. Funct Plant Biol **34:** 550–563

Lytovchenko A, Beleggia R, Schauer N, Isaacson T, Leuendorf JE, Hellmann H, Rose JKC, Fernie AR (2009) Application of GC-MS for the detection of lipophilic compounds in diverse plant tissues. Plant Methods **5:** 11

May P, Christian JO, Kempa S, Walther D (2009) ChlamyCyc: an integrative systems biology database and web-portal for Chlamydomonas reinhardtii. BMC Genomics **10:** 209

Meyer RC, Steinfath M, Lisec J, Becher M, Witucka-Wall H, Torjek O, Fiehn O, Eckardt A, Willmitzer L, Selbig J, et al (2007) The metabolic signature related to high plant growth rate in Arabidopsis thaliana. Proc Natl Acad Sci USA **104:** 4759–4764

Mitchell-Olds T, Pedersen D (1998) The molecular basis of quantitative genetic variation in central and secondary metabolism in Arabidopsis. Genetics **149:** 739–747

Mitchell-Olds T, Schmitt J (2006) Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis. Nature **441:** 947–952

Monforte AJ, Tanksley SD (2000) Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: a tool for gene mapping and gene discovery. Genome **43:** 803–813

Morandini P (2009) Rethinking metabolic control. Plant Sci **176:** 441–451

Morcuende R, Bari RM, Gibon Y, Zheng W, Pant BD, Bläsing O, Usadel B, Czechowski T, Udvardi MK, Stitt M, et al (2007) Genome-wide reprogramming of metabolism, protein synthesis, cellular growth processes and the regulatory networks of Arabidopsis in response to phosphate. Plant Cell Environ **30:** 85–112

Nagel KA, Schurr U, Walter A (2006) Dynamics of root growth stimulation in Nicotiana tabacum in increasing light intensity. Plant Cell Environ **29:** 1936–1945

Neuhaus HE, Kruckeberg AL, Feil R, Gottlieb L, Stitt M (1989) Dosage mutants of phosphoglucose isomerase in the cytosol and chloroplasts of *Clarkia xantiana*. II. Study of the mechanisms which regulate photosynthate partitioning. Planta **178:** 110–122

Neuhaus HE, Quick WP, Siegl G, Stitt M (1990) Control of photosynthate partitioning in spinach leaves: analysis of the interaction between feedforward and feedback-regulation of sucrose synthesis. Planta **181:** 583–592

Nikerel IE, van Winden WA, van Gulik WM, Heijnen JJ (2006) A method for estimation of elasticities in metabolic networks using steady state and dynamic metabolomics data and linlog kinetics. BMC Bioinformatics **7:** 540

Noh K, Gronke K, Luo B, Takors R, Oldiges M, Wiechert W (2006) Metabolic flux analysis at ultra short time scale: isotopically nonstationary C-13 labeling experiments. *In* International Workshop on Molecular Systems Biology. Elsevier Science, Bielefeld, Germany, pp 249–267

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, et al (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet **30:** 190–193

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, et al (2005) The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol **3:** e196

Nordborg M, Weigel D (2008) Next-generation genetics in plants. Nature **456:** 720–723

Okazaki Y, Shimojima M, Sawada Y, Toyooka K, Narisawa T, Mochida K, Tanaka H, Matsuda F, Hirai A, Hirai MY, et al (2009) A chloroplastic UDP-glucose pyrophosphorylase from *Arabidopsis* is the committed enzyme for the first step of sulfolipid biosynthesis. Plant Cell 21: 892–909

Oresic M, Clish C, Davidov EJ, Verheij E, Vogels J, Havekes LM, Neumann E, Adourian A, Naylor S, van der Greef J, et al (2004) Phenotype characterisation using integrated gene transcript, protein and metabolite profiling. Appl Bioinformatics 3: 205–217

Osuna D, Usadel B, Morcuende R, Scheible WR, Gibon Y, Bläsing OE, Thimm O, Höhne M, Günter M, Udvardi MK, et al (2007) Temporal responses of transcripts, enzyme activities and metabolites after adding sucrose to carbon-deprived Arabidopsis seedlings. Plant J 49: 463–491

Park SY, Fung P, Nishimura N, Jensen DR, Fujii H, Zhao Y, Lumba S, Santiago J, Rodrigues A, Chow TFF, et al (2009) Abscisic acid inhibits type 2c protein phosphatases via the pyr/pyl family of start proteins. Science 324: 1068–1071

Piques M, Schulze WX, Höhne M, Usadel B, Gibon Y, Rohwer J, Stitt M (2009) Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in Arabidopsis. Mol Syst Biol 5: 314

Poolman MG, Miguet L, Sweetlove LJ, Fell DA (2009) A genome-scale metabolic model of Arabidopsis and some of its properties. Plant Physiol 151: 1570–1581

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. Am J Hum Genet 67: 170–181

Queitsch C, Sangster TA, Lindquist S (2002) Hsp90 as a capacitor of phenotypic variation. Nature 417: 618–624

Rantalainen M, Cloarec O, Ebbels T, Lundstedt T, Nicholson J, Holmes E, Trygg J (2008) Piecewise multivariate modelling of sequential metabolic profiling data. BMC Bioinformatics 9: 105

Reed JL, Palsson BO (2004) Genome-scale in silico models of E-coli have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. Genome Res 14: 1797–1805

Rischer H, Oresic M, Seppanen-Laakso T, Katajamaa M, Lammertyn F, Ardiles-Diaz W, Van Montagu MC, Inzé D, Oksman-Caldentey KM, Goossens A (2006) Geneto-metabolite networks for terpenoid indole alkaloid biosynthesis in Catharanthus roseus cells. Proc Natl Acad Sci USA 103: 5614–5619

Roessner-Tunali U, Hegemann B, Lytovchenko A, Carrari F, Bruedigam C, Granot D, Fernie AR (2003) Metabolic profiling of transgenic tomato plants overexpressing hexokinase reveals that the influence of hexose phosphorylation diminishes during fruit development. Plant Physiol 133: 84–99

Rolleston FS (1972) A theoretical background to the use of measured intermediates in the study of the control of intermediary metabolism. Curr Top Cell Regul 5: 47–54

Sangster TA, Salathia N, Undurraga S, Milo R, Schelienberg K, Lindquist S, Queitsch C (2008) Hsp90 affects the expression of genetic variation and developmental stability in quantitative traits. Proc Natl Acad Sci USA 105: 2963–2968

Schauer N, Semel Y, Balbo I, Steinfath M, Repsilber D, Selbig J, Pleban T, Zamir D, Fernie AR (2008) Mode of inheritance of primary metabolic traits in tomato. Plant Cell 20: 509–523

Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J, et al (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat Biotechnol 24: 447–454

Scheible WR, Gonzalez-Fontes A, Lauerer M, Muller-Rober B, Caboche M, Stitt M (1997) Nitrate acts as a signal to induce organic acid metabolism and repress starch metabolism in tobacco. Plant Cell 9: 783–798

Scheible WR, Krapp A, Stitt M (2000) Reciprocal diurnal changes of phosphoenolpyruvate carboxylase expression and cytosolic pyruvate kinase, citrate synthase and NADP-isocitrate dehydrogenase expression regulate organic acid metabolism during nitrate assimilation in tobacco leaves. Plant Cell Environ 23: 1155–1167

Scheible WR, Morcuende R, Czechowski T, Osuna D, Fritz C, Palacios-Rojas N, Thimm O, Udvardi MK, Stitt M (2004) Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes and signalling infrastructure by nitrogen. Plant Physiol 136: 2483–2499

Schilling CH, Covert MW, Famili I, Church GM, Edwards JS, Palsson BO (2002) Genome-scale metabolic model of Helicobacter pylori 26695. *In* 9th International Conference on Microbial Genomes. American Society of Microbiology, Lake Arrowhead, CA, pp 4582–4593

Schurr U, Walter A, Rascher U (2006) Functional dynamics of plant growth and photosynthesis: from steady-state to dynamics, from homogeneity to heterogeneity. Plant Cell Environ 29: 340–352

Semel Y, Nissenbaum J, Menda N, Zinder M, Krieger U, Issman N, Pleban T, Lippman Z, Gur A, Zamir D (2006) Overdominant quantitative trait loci for yield and fitness in tomato. Proc Natl Acad Sci USA 103: 12981–12986

Slate J (2005) Quantitative trait locus mapping in natural populations: progress, caveats and future directions. Mol Ecol 14: 363–379

Smilde AK, Jansen JJ, Hoefsloot HJC, Lamers RAN, van der Greef J, Timmerman ME (2005a) ANOVA-Simultaneous Component Analysis (ASCA): a new tool for analyzing designed metabolomics data. Bioinformatics 21: 3043–3048

Smilde AK, van der Werf MJ, Bijlsma S, van der Werff JC, Jellema RH (2005b) Fusion of mass-spectrometry-based metabolomics data. Anal Chem 77: 6729–6736

Smith AM, Stitt M (2007) Coordination of carbon supply and plant growth. Plant Cell Environ 30: 1126–1149

Sønderby IE, Hansen BG, Bjarnholt N, Ticconi C, Halkier BA, Kliebenstein DJ (2007) A systems biology approach identifies a R2R3 myb gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates. PLoS One 2: e1322

Steinfath M, Grith D, Lisec J, Selbig J (2008) Metabolite profile analysis: from raw data to regression and classification. Physiol Plant 132: 150–161

Sterken R, Kiekens R, Coppens E, Vercauteren I, Zabeau M, Inzé D, Flowers J, Vuylsteke M (2009) A population genomics study of the *Arabidopsis* core cell cycle genes shows the signature of natural selection. Plant Cell 21: 2987–2998

Stitt M (1999) The first will be last and the last will be first: non-regulated enzymes call the tune? *In* JA Burrell, MM Bryant, NJ Kruger, eds, Plant Carbohydrate Biochemistry. BIOS Scientific Publishers, Oxford, pp 1–16

Stitt M, Lunn J, Usadel B (2009) Primary photosynthetic metabolism: more than the icing on the cake. Plant J (in press)

Stitt M, Sonnewald U (1995) Regulation of metabolism in transgenic plants. Annu Rev Plant Physiol Plant Mol Biol 46: 341–368

Stitt M, Wirtz W, Heldt HW (1980) Metabolite levels during induction in the chloroplast and extra-chloroplast compartments of spinach protoplasts. Biochim Biophys Acta 593: 85–102

Sulpice R, Pyl ET, Ishihara H, Trenkamp S, Steinfath M, Witucka-Wall H, Gibon Y, Usadel B, Poree F, Piques MC, et al (2009) Starch as a major integrator in the regulation of plant growth. Proc Natl Acad Sci USA 106: 10348–10353

Sweetlove LJ, Fell D, Fernie AR (2008) Getting to grips with the plant metabolic network. Biochem J 409: 27–41

Szymanski J, Bielecka M, Carrari F, Fernie AR, Hoefgen R, Nikiforova VJ (2007) On the processing of metabolic information through metabolite-gene communication networks: an approach for modelling causality. Phytochemistry 68: 2163–2175

Taylor CF, Field D, Sansone SA, Aerts J, Apweiler R, Ashburner M, Ball CA, Binz PA, Bogue M, Booth T, et al (2008) Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. Nat Biotechnol 26: 889–896

Tian D, Araki H, Stahl E, Bergelson J, Kreitman M (2002) Signature of balancing selection in Arabidopsis. Proc Natl Acad Sci USA 99: 11525–11530

Tian F, Stevens NM, Buckler ES (2009) Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. Proc Natl Acad Sci USA 106: 9979–9986

Tiessen A, Hendriks JHM, Stitt M, Branscheid A, Gibon Y, Farre EM, Geigenberger P (2002) Starch synthesis in potato tubers is regulated by post-translational redox modification of ADP-glucose pyrophosphorylase: a novel regulatory mechanism linking starch synthesis to the sucrose supply. Plant Cell 14: 2191–2213

Trygg J, Wold S (2002) Orthogonal projections to latent structures (O-PLS). J Chemometr 16: 119–128

Trygg J, Wold S (2003) O2-PLS, a two-block (X-Y) latent variable regression (LVR) method with an integral OSC filter. J Chemometr 17: 53–64

Turgeon R, Wolf S (2009) Phloem transport: Cellular pathways and molecular trafficking. Annu Rev Plant Biol 60: 207–221

Walter A, Silk WK, Schurr U (2009) Environmental effects on spatial and temporal patterns of leaf and root growth. Annu Rev Plant Biol **60:** 279–304

Warner JR (1999) The economics of ribosome biosynthesis in yeast. Trends Biochem Sci **24:** 437–440

Watanabe M, Mochida K, Kato T, Tabata S, Yoshimoto N, Noji M, Saito K (2008) Comparative genomics and reverse genetics analysis reveal indispensable functions of the serine acetyltransferase gene family in *Arabidopsis*. Plant Cell **20:** 2484–2496

Weigel D, Mott R (2009) The 1001 Genomes project for Arabidopsis thaliana. Genome Biol **10:** 107

Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler ES (2004) Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell **16:** 2719–2733

Yonekura-Sakakibara K, Tohge T, Matsuda F, Nakabayashi R, Takayama H, Niida R, Watanabe-Takahashi A, Inoue E, Saito K (2008) Compre-hensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in *Arabidopsis*. Plant Cell **20:** 2160–2176

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics **178:** 539–551

Yuan J, Doucette CD, Fowler WU, Feng XJ, Piazza M, Rabitz HA, Wingreen NS, Rabinowitz JD (2009) Metabolomics-driven quantitative analysis of ammonia assimilation in E. coli. Mol Syst Biol **5:** 302

Zhao Y, Chow TF, Puckrin RS, Alfred SE, Korir AK, Larive CK, Cutler SR (2007) Chemical genetic interrogation of natural variation uncovers a molecule that is glycoactivated. Nat Chem Biol **3:** 716–721

Zou F, Gelfond JA, Airey DC, Lu L, Manly KF, Williams RW, Threadgill DW (2005) Quantitative trait locus analysis using recombinant inbred intercrosses: theoretical and empirical considerations. Genetics **170:** 1299–1311